Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 1428–1432, March 1982
Biochemistry

Structure of the DNA-binding region of *lac* repressor inferred from its homology with *cro* repressor

(gene regulation/DNA–protein interaction/amino acid sequence/protein structure)

B. W. MATTHEWS*, D. H. OHLENDORF*, W. F. ANDERSON†, AND Y. TAKEDA‡

*Institute of Molecular Biology and Department of Physics, University of Oregon, Eugene, Oregon 97403; †Medical Research Council Group on Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 247, Canada; and ‡Chemistry Department, University of Maryland, Baltimore County, Maryland 21228

Communicated by Peter H. von Hippel, November 9, 1981

ABSTRACT **It is shown that the amino acid sequence and the DNA gene sequence of the 25 amino-terminal residues of the *lac* repressor protein of *Escherichia coli* are homologous with the sequences of five DNA-binding proteins: the *cro* repressor proteins from phage λ and phage 434, the *c*I and *c*II proteins from phage λ, and the repressor protein from *Salmonella* phage P22. The region of homology between *lac* repressor and the other proteins coincides with the principal DNA-binding region of *cro* repressor. In particular, residues Tyr-17 through Gln-26 of *lac* repressor correspond to the α-helix Gln-27 through Ala-36 of *cro* repressor, which we have postulated to bind within the major groove of the DNA and to be primarily responsible for the recognition of the DNA operator region by the protein [Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. (1981) *Nature (London)* 290, 754–758]. By analogy with *cro* repressor, we propose that residues 17–26 of *lac* repressor are α-helical and that this helix and a twofold-related α-helix in an adjacent subunit bind within successive major grooves of the *lac* operator, which is in a right-handed Watson–Crick B-DNA conformation. Also, by analogy with *cro* repressor, we suggest that residues Thr-5 through Ala-13 of *lac* repressor form a second α-helix and contribute, in part, to DNA binding. The proposed structure for the DNA-binding region of *lac* repressor is consistent with chemical protection data and with genetic experiments identifying the probable locations of a number of the residues of the repressor protein that either do or do not participate in DNA binding.**

The way in which certain proteins recognize specific base sequences within double-stranded DNA has been a central problem in molecular biology. From a biochemical and genetic point of view the best-characterized example of such an interaction is the *lac* repressor–*lac* operator system of *Escherichia coli* (1–3), but the lack of good crystals has prevented the determination of the three-dimensional structure of the *lac* repressor protein. Recently, we determined the structure of the *cro* repressor ("cro") from bacteriophage λ and proposed a model for its interaction with DNA (4). We have also shown on the basis of amino acid sequence and DNA gene sequence comparisons that the helical DNA-binding domain of cro probably occurs in a number of other DNA-binding proteins, including *cro* protein from bacteriophage 434, *c*I and *c*II proteins from bacteriophage λ, and the *Salmonella* phage P22 repressor protein (5). Here we suggest that the DNA-binding region of *lac* repressor also includes an α-helical region that interacts with its DNA operator in a manner similar to that proposed for cro. The suggested mode of interaction of *lac* repressor with *lac* operator is consistent with a variety of biochemical and genetic evidence and is also strikingly similar to that proposed previously by Adler *et al.* (6).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Comparison of DNA-binding proteins

Fig. 1 shows a comparison of the amino-terminal sequences of a series of proteins that bind to sequence-specific regions of double-stranded DNA (5). cro and 434-cro are small repressor proteins from bacteriophage λ and bacteriophage 434, respectively (10–13). "cI" (often referred to as λ repressor) and "P22" are larger repressor proteins from phage λ and from *Salmonella* phage P22, respectively, that, under different circumstances, can mediate positive or negative control of gene transcription (14–18). "cII" is also larger than cro and, in conjunction with another protein (cIII), acts as a positive regulator of transcription in bacteriophage λ (16, 17, 19, 20). With the exception of cro and cI, these five proteins all recognize different sequences on the DNA.

As can be seen in Fig. 1, there is extensive amino acid sequence homology between the five DNA-binding proteins. The correspondence between the five proteins can also be seen in the DNA gene sequences that code for the respective polypeptides, and, on the basis of this sequence homology, we have argued (5) that these proteins all have in common a region of tertiary structure corresponding to the segments labeled α_1, α_2, and α_3. In the cro structure, α_1 and α_2 are "structural" α-helices, whereas α_3 is the "DNA recognition" α-helix, which we have postulated to lie within the major groove of B-form DNA and to be primarily responsible for the specific recognition of the DNA by the protein (4).

A series of comparisons of both the amino acid sequence (21, 22) and the DNA coding sequence (23) indicates that the *lac* repressor protein from *E. coli* ("lac") may also have structural features in common with the other DNA-binding proteins. In Fig. 1 we have included the first 38 amino acids of lac, aligned to maximize the homology with the other proteins. The homology is most striking in the region 19–32 of cro (9–22 of lac), where Gly-24 is invariant and Ala-20 and Val-25 of cro, which are conserved in four of the five proteins, also occur in lac.

The homology between lac and the other proteins can also be seen at the level of the DNA sequences that code for the respective polypeptides. Table 1 summarizes comparisons of the DNA coding sequences for the 60 bases (i.e., 20 amino acids) corresponding to Met-12 to Asn-31 of cro (or Met-1 to Ser-21 of lac). Each entry in the table gives the fraction of the bases that are common for a given pair of proteins. If the respective DNA sequences were unrelated, then this ratio would be expected to equal approximately 0.25 (one base in four in common; each of the four bases occurs with approximately equal frequency in each of the gene sequences being compared). As can be seen, every entry in the table exceeds the random value, and in some cases more than 50% of the bases are the same. The

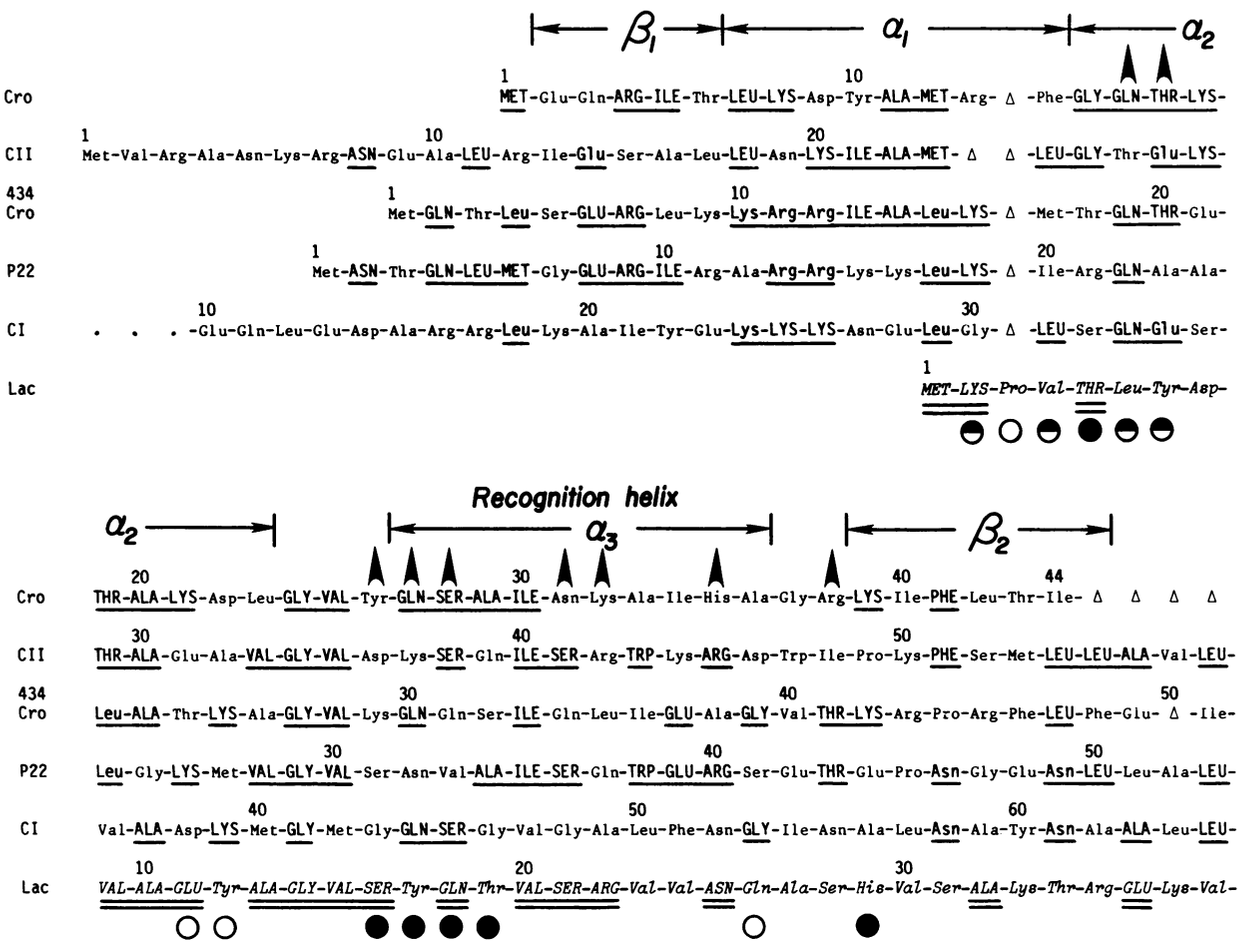

FIG. 1. Comparison of the amino-terminal amino acid sequence of *lac* repressor with five other DNA-binding proteins. Residues that are homologous within the five DNA-binding proteins are indicated by different typeface and a single underline. Residues of *lac* repressor that are common to one or more of the five proteins are capitalized and have a double underline. The letters α and β show the locations of the α-helices and β-sheet strands of *cro* protein. Residues of cro that are presumed, from model building, to interact with DNA are capped by an arrowhead. Δ indicates an assumed deletion. The solid circles underneath the lac sequence indicate those locations where known mutations dramatically reduce DNA-binding ability but do not interfere with inducer binding (i.e., they do not simply destabilize the whole protein). (Phenotypically similar sites of substitution also occur at positions 45, 47, and 54 but cannot be shown in Fig. 1.) The half-filled circles indicate locations where amino acid substitutions may reduce DNA-binding affinity or where the reduction in binding is weak. Open circles indicate sites where substitution does not interfere with DNA binding. (Data on mutants taken from refs. 2, 3, and 6–9).

table also includes the average value for each protein compared, in turn, with the other five. Somewhat surprisingly, the average agreement ratio for lac is the second highest in the series.

Significance of the sequence homology

It might be asked whether the amino acid homology between *lac* repressor and the other proteins listed in Fig. 1 is sufficiently good to provide convincing evidence that *lac* repressor is, in fact, related to the other proteins. In estimating the significance of the homology, the essential factor that has to be taken into account is that the first five sequences are already aligned among themselves so that the *lac* sequence is aligned *simultaneously* with each of the other five sequences. It is this simultaneous correspondence that increases the overall significance of the homology. For example, if one compares the first 22 residues of *lac* repressor with *cro* repressor alone, there are 4 equivalences (Fig. 1). A frequency of 4 in 22 is above the chance value of about 1 in 15 for a typical amino acid composition (24) but, on its own, is not unusual (significance of 2.2σ). However, in the case of simultaneous agreement between one protein and a number of others, it is necessary to sum all the equivalences between the one protein and each of the other proteins (24, 25). For the first 22 residues of *lac* repressor there are 110 (i.e., 22×5) possible matches with the other proteins, and, of these, 27 residues are identical (Fig. 1). An agreement of 27 amino acid residues out of 110 (significance of 7.5σ) is very unlikely to have occurred by chance and is usually taken as clear-cut evidence for significant sequence homology (24).

In addition to using the amino acid sequence homology, we also attempted to evaluate the significance of the homology between the gene sequences of *lac* repressor and the other proteins. To see whether the agreement ratios between lac and the other proteins listed in Table 1 are better than for other alignments, we took the first 21 amino acids of lac, translocated them to the left or the right of the alignment shown in Fig. 1, and calculated the number of corresponding bases. Table 1 summarizes the results for the ten cases in which the gene sequence beginning at Met-1 of lac was aligned, in turn, with gene sequences beginning at Leu-7 through Gln-17 of cro (and the corresponding alignments for the other proteins). In each case, the average base agreement ratio for the 10 possible alignments is close to the expected value of 0.25. The best agreement ratios for the 10 possible alignments are, with one exception (lac vs. cII), substantially poorer than the ratios obtained for the alignment of lac as in Fig. 1. The high agreement ratios for lac vs. 434-cro and lac vs. P22 repressor (52% and 45% of bases in com-

Table 1. DNA sequence comparisons for the genes of six DNA-binding proteins*

	cro	cII	434-cro	P22	cI	lac
Sequences aligned as in Fig. 1						
cro	—	0.60	0.43	0.42	0.32	0.40
cII	0.60	—	0.28	0.35	0.35	0.37
434-cro	0.43	0.28	—	0.45	0.35	0.52
P22	0.42	0.35	0.45	—	0.37	0.45
cI	0.32	0.35	0.35	0.37	—	0.38
lac	0.40	0.37	0.52	0.45	0.38	—
Average ratio	0.43	0.39	0.41	0.41	0.35	0.42
lac sequence aligned in 10 different ways (see text)						
Average of 10 ratios	0.24	0.29	0.24	0.25	0.25	—
Highest ratio	0.35	0.41	0.35	0.35	0.30	—

* The comparisons are made for the 60 bases corresponding to Met-1 to Ser-21 of *lac* repressor. Each entry gives the fraction of the 60 bases that are common for a given pair of proteins.

mon) relative to the corresponding highest control values (about 35%) are particularly noteworthy. Furthermore, it has to be emphasized that the "average ratios" in Table 1 were obtained for *lac* repressor aligned simultaneously with the five other DNA-binding proteins, whereas the "highest ratios" were selected from the 10 possible alignments of lac with each of the individual proteins in turn.

In another test we took the gene sequence for the whole *lac* repressor molecule and calculated the base agreement ratio as defined in Table 1 for every possible 60-base segment of lac aligned with the other five proteins. For these 1018 alignments, the average fraction of the 300 (i.e., 60 × 5) possible bases in common between lac and the other five proteins was 0.249 (compare the expected value of 0.25), the standard deviation, σ, was 0.036 and the highest value of 0.424 was for the alignment of the lac sequence shown in Fig. 1. The significance of the highest value is 4.9σ. Thus, this method of evaluating the significance of the sequence homology, which is essentially that of Fitch (26), also suggests that the proposed homology between *lac* repressor and the other proteins is very unlikely to be due to chance.

Discussion

Taken together, the amino acid homology and the gene sequence homology between *lac* repressor and the other five DNA-binding proteins strongly suggest that at least a part of *lac* repressor evolved from the same precursor as did the other proteins. Because of the apparent sequence homology, we postulate that the structure of the amino terminus of *lac* repressor is similar to that of residues 11–40 of cro. In particular, we suggest that residues Tyr-17 to Gln-26 of lac form an α-helix that binds in the major groove of right-handed B-form DNA and is primarily responsible for the recognition of the specific operator sequence by the repressor. The proposed conformation of the 30 amino-terminal residues of *lac* repressor and their interaction with operator DNA are shown in Figs. 2 and 3.

Convincing sequence homology between two proteins provides compelling evidence that they have similar three-dimensional structures (30), but it has to be noted that the apparent amino acid and DNA sequence correspondence for lac is restricted to a stretch of about 25 amino acid residues, so that the significance of the implied structural homology might be questioned. Clearly one has to ask if the postulated structural correspondence between lac and the other proteins (cro in particular) is consistent with the known properties of *lac* repressor.

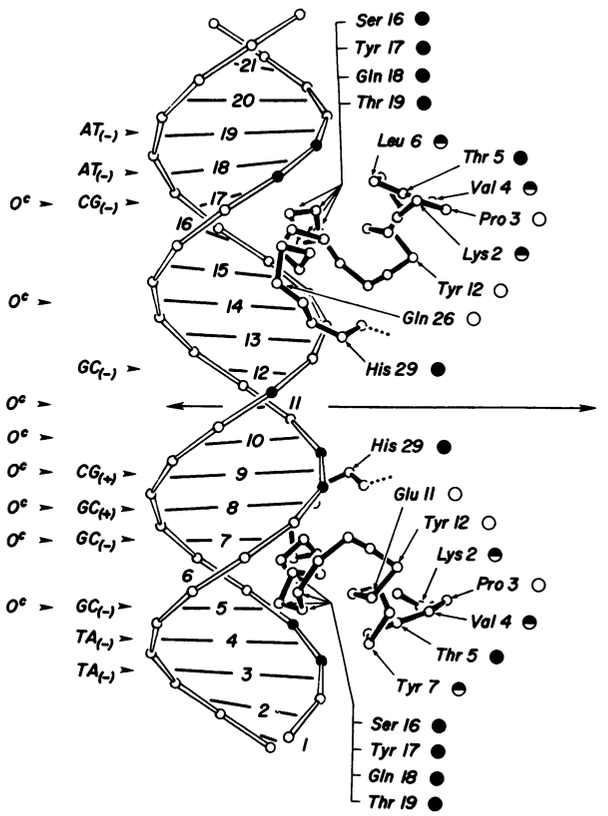

FIG. 2. Diagram showing the proposed interaction of a pair of two-fold-related DNA-binding regions of *lac* repressor with operator DNA. The twofold axis that relates the subunits of the repressor also passes through the central base pair (no. 11) of the operator. The backbone of the 30 amino-terminal residues of *lac* repressor (shown solid) is drawn as if it were identical with the backbone of residues 11–40 of *cro* repressor, although it is likely that the 5 or 6 amino-terminal residues of lac interact more closely with the DNA than is suggested by the figure (see text). Solid circles and half-solid circles indicate residues that are thought, on the basis of known mutations at these locations, to potentially interact with the DNA. Open circles show residues that are presumed, on the basis of genetic evidence, not to interact with the DNA (see caption to Fig. 1). Base pairs given at the left are those for which binding of the repressor modifies the availability of the purine to methylation by dimethyl sulfate (27). G·C or C·G methylation occurs in the major groove, whereas A·T or T·A methylation occurs in the minor groove. A minus indicates that *lac* repressor binding inhibits methylation; a plus indicates enhanced methylation in the presence of repressor. Also shown at the left are the locations of known O^c operator mutations, which reduce the affinity of the repressor for the operator (data from ref. 28). The phosphates of the DNA drawn solid are those for which alkylation with ethylnitrosourea interferes with repressor binding (data from W. Gilbert and A. Maxam, reported in ref. 29).

Both biochemical and genetic evidence shows that most, if not all, of the DNA-binding region of lac resides within the 60 amino-terminal residues of the protein (2, 3, 6, 31, 32). Under certain conditions, tryptic cleavage of the intact tetrameric repressor yields four amino-terminal "headpieces" of 51 or 59 residues and a tetrameric "core" with full inducer binding activity (31, 32). These headpieces bind nucleic acids, interact with the *lac* operator (33), and protect the same set of bases against methylation as does the intact repressor (34).

Predictions of the secondary structure of *lac* repressor give conflicting results (6–8, 35–38), some of which are in agreement with the proposed conformation.

A series of genetic analyses has pinpointed a number of residues of lac that may be directly involved with DNA binding

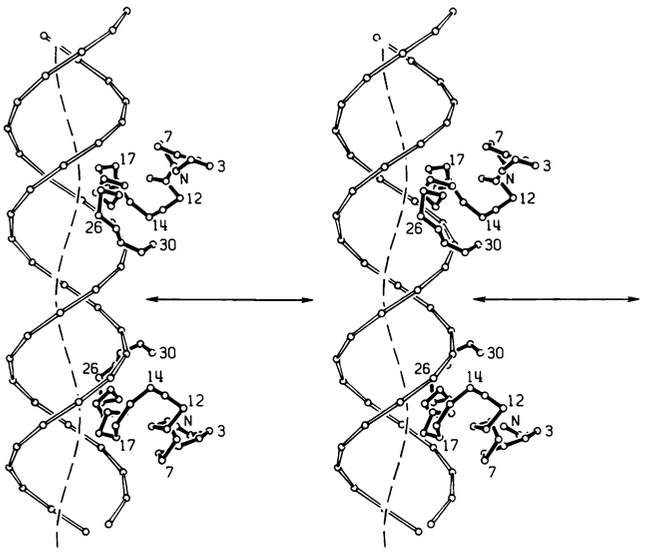

FIG. 3. Stereo drawing showing the proposed conformation of the 30 amino-terminal (N) residues of *lac* repressor (drawn solid) and their interaction with operator DNA (cf. Fig. 2). The horizontal line shows the location of the twofold axis that passes through the center of the operator and also relates the repressor subunits. The broken line extending along the DNA shows the location of the bottom of the major groove.

(2, 3, 6–9). In Fig. 1 the solid and half-filled circles below the lac sequence indicate residues that probably interact directly with the DNA or whose substitution modulates DNA binding. These sites can be compared with the locations of the residues of cro that we believe, from model building, to interact with DNA. [The model-building experiments are based on a model of cro partially refined to a crystallographic residual of 27% at 2.2-Å resolution (unpublished results).] The locations of the putative DNA-binding residues are also shown in Figs. 2 and 3. As can be seen, each of these residues is in a position where it could interact with the DNA, especially if one makes the reasonable assumption that the five or six amino-terminal residues undergo a slight conformational adjustment so as to lie against the DNA. Residues Ser-16 to Thr-19 of lac, which were presumed on the basis of the genetic studies to be critically involved in DNA binding, are located at the beginning of the postulated "recognition helix" (Figs. 1–3). Residues Thr-5 to Tyr-7 of lac are aligned with the DNA-binding region Gly-15 to Thr-17 of cro. Also, His-29 of lac closely corresponds to Arg-38 to Lys-39 of cro. Residues Glu-11 and Tyr-12 of lac, which were thought not to be directly involved with DNA binding, are located away from the DNA. Gln-26 of lac is at the end of the presumed recognition helix but is directed away from the DNA.

In 1972, Müller-Hill and coworkers (6) proposed that the DNA-binding region of lac consisted of a protruding α-helix that extended from amino acids 17 to 33. Our analysis suggests that, in the main, these workers (6) were very close to the truth, even though their ideas concerning the role of the four subunits of lac need modification. The structure we propose for the DNA-binding region of *lac* repressor is not consistent with models in which the operator-binding region adopts an extended β-strand conformation (35, 39) or a left-handed δ-helix (40, 41). However, as can be seen in Fig. 3, the proposed structure is consistent with the recent nuclear magnetic resonance study of Arndt *et al.* (42) indicating that Tyr-7 and Tyr-17 are very close to each other, possibly ring stacked, and that Tyr-12 and His-29 are at or near the surface of the protein.

On the basis of the knowledge of the cro structure we suggest that the *lac* repressor binds to its operator with a twofold symmetry axis of the repressor coincident with the twofold axis of the DNA operator (4, 43). Probably one pair of twofold-related subunits of the repressor tetramer makes contact with one operator (43–48) and, in addition, the second pair of twofold-related subunits constitutes a second operator-binding site, explaining the observed stoichiometry of two operators bound per repressor tetramer (49, 50). We postulate that a pair of twofold-related α-helices lies within successive major grooves of the DNA, as is the case for the cro protein (4). Recognition of a specific DNA sequence is mediated primarily by a complementary set of hydrogen bonds between the side chains of the protein and the parts of the base pairs that are accessible in the major grooves of the DNA. The action of inducer in reducing the affinity of *lac* repressor for its operator DNA is most easily envisaged as causing a slight conformational rearrangement of the subunits so that the relative alignment of the two DNA-binding helices is perturbed. This could be a general phenomenon for allosterically regulated DNA-binding proteins. The separation of the DNA-binding function into one domain and the regulatory function into a second domain also seems to be generally true for such proteins.

On the basis of the model-building experiments with cro, the 30 or so amino-terminal residues of lac would make interactions predominantly on one face of the DNA extending to seven or eight base pairs on either side of the central dyad. This is consistent with the locations of recognition sites of *lac* operator inferred from chemical protection (27, 51), genetic (28), and modification (52) experiments. The proposed structure of *lac* repressor as drawn in Figs. 2 and 3 would not explain the outermost methylation protection/modification sites, which are seven and eight base pairs from the center of the operator and located in the minor groove of the DNA. However, we have suggested above that the five or six amino-terminal residues of lac undergo a conformational adjustment and probably interact more closely with the DNA than is suggested by Figs. 2 and 3. These interactions might extend, for the respective lac subunits, to base pairs 1 and 21 in Fig. 2 and could well account for the outermost protection/modification sites. Another possibility is that the part of the lac headpiece including residues 45 and 47, which are possibly involved with DNA binding, interacts with the outermost part of the operator. It is also possible that the region of the headpiece including Lys-33, Arg-35, and Lys-37 interacts with the central part of the operator (Figs. 2 and 3). In our model for the lac headpiece structure, Lys-2, Arg-22, and His-29 are in a position to make contact with the DNA, and we have suggested that some or all of Lys-33, Arg-35, Lys-37, and the amino terminus of the protein could interact as well. Allowing for the combined interactions in the twofold-related subunits accounts for the experimental estimation that the nonspecific interaction of *lac* repressor with DNA involves about 11 ionic interactions with the sugar phosphate backbone and also that binding of the repressor is accompanied by uptake of two protons, probably on histidines or α-amino groups (53–56).

Although the proposed model for the DNA-binding region of *lac* repressor may not include all parts of the protein that interact with the DNA, it could well account for most, if not all, of the sequence-specific interactions. As such, it can be tested by matching models of the protein with native, mutant, and modified operator sequences. Detailed model building can be used to try to understand the structural basis for the sequence specificity of the recognition of operator DNA by the *cro* repressor. Similarly, the proposed structure for the DNA-binding region of *lac* repressor will permit the building of a detailed model of a putative *lac* repressor·*lac* operator complex. The consistency of this model with the wealth of experimental data

on the *lac* system should serve as a test for the validity of the structure that we have proposed.

We thank Dr. Peter H. von Hippel for helpful discussions and comments on the manuscript. This work was supported in part by grants from the National Science Foundation (PCM-8014311), the National Institutes of Health (GM 20066; GM 28138), the Medical Research Council Group on Protein Structure and Function of Canada, and the M. J. Murdock Charitable Trust. D.H.O. was supported in part by a Damon Runyon–Walter Winchell Cancer Fund Postdoctoral Fellowship (DRG-303-F).

1. Jacob, F. & Monod, J. (1961) *J. Mol. Biol.* **3**, 318–356.
2. Sund, H. & Blauer, G., eds. (1975) *Protein-Ligand Interactions* (de Gruyter, Berlin).
3. Miller, J. H. & Reznikoff, W. S., eds. (1978) *The Operon* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
4. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. (1981) *Nature (London)* **290**, 754–758.
5. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. *J. Mol. Biol.*, in press.
6. Adler, K., Beyreuther, K., Fanning, E., Geisler, N., Gronenborn, B., Klemm, A., Müller-Hill, B., Pfahl, M. & Schmitz, A. (1972) *Nature (London)* **237**, 322–327.
7. Müller-Hill, B. (1975) *Prog. Biophys. Mol. Biol.* **30**, 227–252.
8. Weber, K., Platt, T., Ganem, D. & Miller, J. H. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 3624–3628.
9. Miller, J. H., Coulondre, C., Hofer, M., Schmeissner, U., Sommer, H. & Schmitz, A. (1979) *J. Mol. Biol.* **131**, 191–222.
10. Folkmanis, A., Takeda, Y., Simuth, J., Gussin, G. & Echols, H. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2249–2253.
11. Takeda, Y., Folkmanis, A. & Echols, H. (1977) *J. Biol. Chem.* **252**, 6177–6183.
12. Hsiang, M. W., Cole, R. D., Takeda, Y. & Echols, H. (1977) *Nature (London)* **270**, 274–275.
13. Grosschedl, R. & Schwarz, E. (1979) *Nucleic Acids Res.* **6**, 867–881.
14. Sauer, R. T. & Anderegg, R. (1978) *Biochemistry* **17**, 1092–1100.
15. Ptashne, M., Jeffrey, A., Johnson, A. D., Maurer, R., Meyer, B. J., Pabo, C. O., Roberts, T. M. & Sauer, R. T. (1980) *Cell* **19**, 1–11.
16. Herskowitz, I. & Hagen, D. A. (1980) *Annu. Rev. Genet.* **14**, 399–445.
17. Echols, H. (1980) in *The Molecular Genetics of Development*, eds. Loomis, W. & Leighton, T. (Academic, New York), pp. 1–16.
18. Sauer, R. T., Pan, J., Hopper, P., Hehis, K., Brown, J. & Poteete, A. R. (1981) *Biochemistry* **20**, 3591–3598.
19. Schwarz, E., Scherer, G., Hobom, G. & Kossel, H. (1978) *Nature (London)* **272**, 410–414.
20. Shimatake, H. & Rosenberg, M. (1981) *Nature (London)* **292**, 128–132.
21. Beyreuther, K., Adler, K., Geisler, N. & Klemm, A. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 3576–3580.
22. Beyreuther, K. (1978) *Nature (London)* **274**, 767.
23. Farabaugh, P. J. (1978) *Nature (London)* **274**, 765–769.
24. Doolittle, R. F. (1981) *Science* **214**, 149–159.
25. Jukes, T. H. & Cantor, C. R. (1969) in *Mammalian Protein Metabolism*, ed. Munro, H. N. (Academic, New York), Vol. 3, pp. 21–132.
26. Fitch, W. M. (1966) *J. Mol. Biol.* **16**, 9–16.
27. Ogata, R. & Gilbert, W. (1979) *J. Mol. Biol.* **132**, 709–728.
28. Gilbert, W., Gralla, J., Majors, J. & Maxam, A. (1978) in *The Operon*, eds. Miller, J. H. & Reznikoff, W. S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 193–210.
29. Barkley, M. D. & Bourgeois, S. (1978) in *The Operon*, eds. Miller, J. H. & Reznikoff, W. S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 177–220.
30. Matthews, B. W. (1977) in *The Proteins*, eds. Neurath, H. & Hill, R. L. (Academic, New York), Vol. 3, 3rd Ed., pp. 403–590.
31. Platt, T., Files, J. G. & Weber, K. (1973) *J. Biol. Chem.* **248**, 110–121.
32. Files, J. G. & Weber, K. (1976) *J. Biol. Chem.* **251**, 3386–3391.
33. Geisler, N. & Weber, K. (1977) *Biochemistry* **16**, 938–943.
34. Ogata, R. T. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5851–5854.
35. Chou, P. Y., Adler, A. J. & Fasman, G. D. (1975) *J. Mol. Biol.* **96**, 29–45.
36. Bourgeois, S., Jernigan, R. L., Shousun, C. S., Kabat, E. A. & Wu, T. T. (1979) *Biopolymers* **18**, 2625–2643.
37. Lim, V. I. (1974) *J. Mol. Biol.* **88**, 857–872.
38. Lim, V. I. (1974) *J. Mol. Biol.* **88**, 873–894.
39. Gursky, G. V., Tumanyan, V. G., Zasedatelev, A. S., Zhuse, A. L., Grokhovsky, S. L. & Gottikh, B. P. (1977) in *Nucleic Acid-Protein Recognition*, ed. Vogel, H. J. (Academic, New York), pp. 189–217.
40. Chandrasekaran, R., Jardetzky, T. S. & Jardetzky, O. (1979) *FEBS Lett.* **101**, 11–14.
41. Riberio, A. A., Wemmer, D., Bray, R. P., Wade-Jardetzky, N. G. & Jardetzky, O. (1981) *Biochemistry* **20**, 818–823.
42. Arndt, K. T., Boschelli, F., Lu, P. & Miller, J. H. (1981) *Biochemistry* **20**, 6109–6118.
43. Steitz, T. A., Richmond, T. J., Wise, D. & Engelman, D. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 593–597.
44. Müller-Hill, B., Gronenborn, B., Kania, J., Schlotmann, M. & Beyreuther, K. (1977) in *Nucleic Acid-Protein Recognition*, ed. Vogel, H. J. (Academic, New York), pp. 219–236.
45. Charlier, M., Maurizot, J. C. & Zaccai, G. (1980) *Nature (London)* **286**, 423–425.
46. Geisler, N. & Weber, K. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 3103–3106.
47. Kania, J. & Brown, T. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 3529–3533.
48. McKay, D. B., Pickover, C. A. & Steitz, T. A., *J. Mol. Biol.*, in press.
49. Culard, F. & Maurizot, J. C. (1979) *Nucleic Acids Res.* **9**, 5175–5184.
50. Dunaway, M. & Matthews, K. S. (1980) *J. Biol. Chem.* **255**, 10120–10127.
51. Ogata, R. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5851–5854.
52. Goeddel, D. V., Yansura, D. G. & Caruthers, M. H. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3578–3582.
53. Revzin, A. & von Hippel, P. H. (1977) *Biochemistry* **16**, 4769–4776.
54. de Haseth, P. L., Gross, C. A., Burgess, R. R. & Record, M. T., Jr. (1977) *Biochemistry* **16**, 4777–4783.
55. de Haseth, P. L., Lohman, T. M. & Record, M. T., Jr. (1977) *Biochemistry* **16**, 4783–4790.
56. Record, M. T., Jr., de Haseth, P. L. & Lohman, T. M. (1977) *Biochemistry* **16**, 4791–4796.